

Bootstrap

José R. Berrendero

- Introducción
- La función de distribución empírica
- Bootstrap: idea básica
- Consistencia
- ¿Cuándo falla el bootstrap?
- Intervalos de confianza
- Referencias

Introducción

Los métodos de remuestreo son una herramienta indispensable para la ciencia de datos. Son métodos que involucran extraer repetidamente muestras de los datos de entrenamiento con el fin de obtener información de las propiedades de un estimador o del ajuste de un modelo. En este curso, aparecerán sobre todo dos métodos de remuestreo: el bootstrap y la validación cruzada. En este tema estudiaremos los principios fundamentales del bootstrap, mientras que en el tema siguiente estudiaremos una de las aplicaciones de la validación cruzada.

Un problema central en estadística es aproximar la distribución en el muestreo de un estadístico, esto es, conocer cómo se distribuyen los valores que toma al muestrear repetidamente de la población. Esta distribución permite valorar la precisión de un estimador (su error típico) y es la base de los principales procedimientos de inferencia. Con el fin de abordar este problema de una forma muy flexible, Bradley Efron (https://en.wikipedia.org/wiki/Bradley_Efron) introdujo en 1979 los métodos bootstrap, en los que se combinan técnicas de simulación con un principio general según el cual cualquier cantidad desconocida que dependa de una distribución F se puede aproximar reemplazando F por un estimador adecuado \widehat{F} obtenido a partir de los datos.

La palabra *bootstrap* alude a una de las aventuras del Barón de Münchhausen, escritas en el siglo XVIII por R. E. Raspe, según la cual el Barón cayó a las aguas de un profundo lago y consiguió salir tirando de los cordones de sus botas (de donde procede la expresión en inglés *to pull oneself up by one's own bootstrap*).

La función de distribución empírica

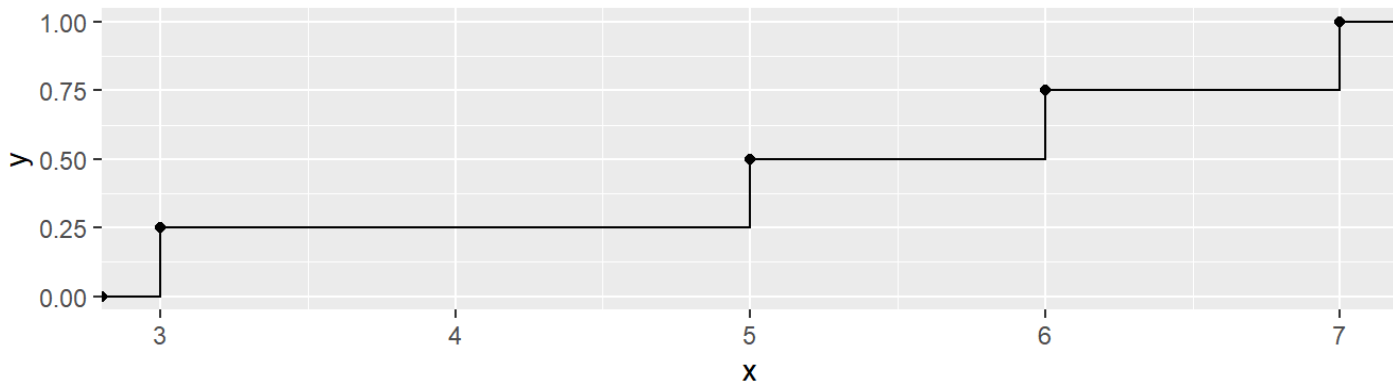
Antes de introducir el bootstrap es necesario conocer la función de distribución empírica asociada a una muestra.

Dada una muestra de n individuos X_1, \dots, X_n , se define la correspondiente **función de distribución empírica** como

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}},$$

donde \mathbb{I}_A vale 1 si el suceso A ocurre, y 0 si A no ocurre. Antes de observar la muestra, $F_n(x)$ es una variable aleatoria. Para cada realización de la muestra tendremos una realización de la función de distribución empírica. Por ejemplo, si la muestra es $x_1 = 3, x_2 = 5, x_3 = 6, x_4 = 7$, la función de distribución empírica se ha representado en la figura siguiente:

```
# Cálculo y representación de la f. de dist. empírica
df <- data.frame(x = c(3, 5, 6, 7))
ggplot(df) +
  stat_ecdf(aes(x)) +
  stat_ecdf(aes(x), geom = 'point')
```



Cuestiones

1. F_n es monótona no decreciente, continua por la derecha, $\lim_{n \rightarrow \infty} F_n(x) = 1$ y $\lim_{n \rightarrow -\infty} F_n(x) = 0$. Como consecuencia, F_n es una auténtica función de distribución. ¿Cómo se reparte la probabilidad según esta distribución?
2. ¿Como se podría obtener una muestra de v.a.i.i.d. X_1^*, \dots, X_n^* procedente de la distribución F_n ?
3. Si X^* tiene distribución dada por F_n , ¿cuánto vale $E_{F_n}(X^*)$?
4. Fijamos x , ¿cuál es la distribución de la variable aleatoria $nF_n(x)$?
5. Como consecuencia de la respuesta a la cuestión anterior determina el valor de $E[F_n(x)]$ y $\text{Var}[F_n(x)]$.

Dado $\epsilon > 0$, usando la desigualdad de Chebychev vemos que

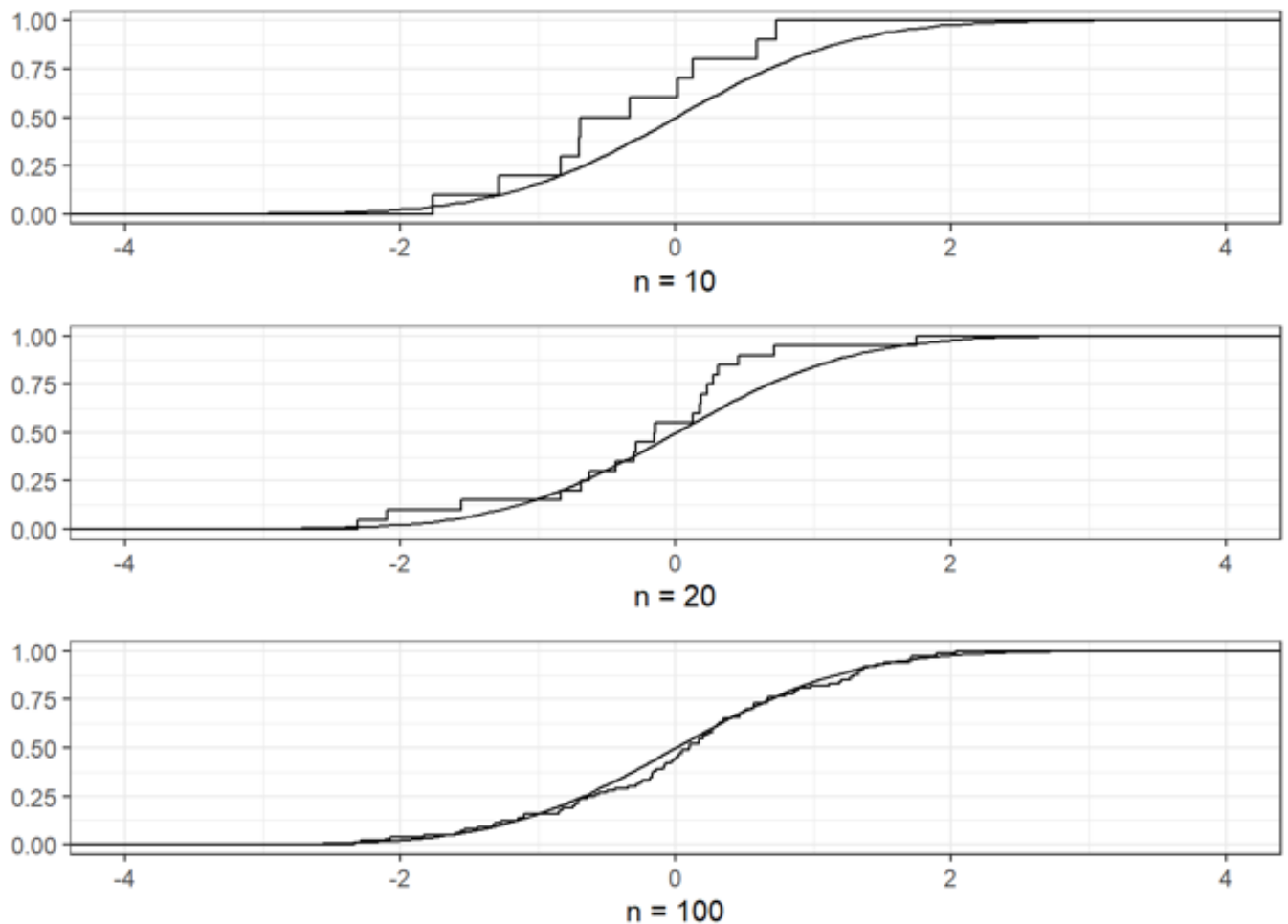
$$P(|F_n(x) - F(x)| > \epsilon) \leq \frac{\text{Var}[F_n(x)]}{\epsilon^2} \rightarrow 0.$$

Como consecuencia, para todo $x \in \mathbb{R}$, $F_n(x) \rightarrow_p F(x)$. De hecho, se verifica un resultado mucho más fuerte:

Teorema de Glivenko-Cantelli. Sea F_n la función de distribución empírica correspondiente a una muestra de v.a.i.i.d. X_1, \dots, X_n con distribución F . Entonces,

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0, \text{ c.s.}$$

A medida que tenemos más datos procedentes de F , la función de distribución empírica se aproxima uniformemente a F con probabilidad uno. Esto significa que con un número suficientemente grande de datos podemos reconstruir la verdadera distribución F con el grado de precisión que queramos (véase la figura siguiente).



La cota de Dvoretzky, Kiefer y Wolfowitz da una información más precisa sobre la diferencia entre la distribución empírica y la verdadera:

Cota de Dvoretzky, Kiefer y Wolfowitz. Existe una constante $C > 0$, independiente de F , tal que, para todo $\epsilon > 0$,

$$P(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > \epsilon) \leq C e^{-2n\epsilon^2}.$$

De hecho, Massart en 1990 probó que si $n\epsilon^2 > 0.5 \log 2$, entonces podemos tomar $C = 2$.

Un método general de estimación en estadística es el principio de sustitución o *plug-in* que consiste en estimar cualquier cantidad que dependa de F , por ejemplo $\theta = g(F)$ por la análoga que resulta de sustituir F por F_n , $\hat{\theta} = g(F_n)$. Esta idea tan simple es esencialmente la que está detrás del método de momentos, es muy potente y cuando se lleva a sus últimas consecuencias conduce a los *métodos de remuestreo bootstrap* ([https://es.wikipedia.org/wiki/Bootstrapping_\(estadística\)](https://es.wikipedia.org/wiki/Bootstrapping_(estadística))) que vamos a estudiar a continuación.

Una aplicación: el contraste de Kolmogorov-Smirnov

Tenemos una muestra X_1, \dots, X_n procedente de una distribución F y queremos contrastar la hipótesis nula de que F coincide con una distribución **continua** dada $F_0(H_0 : F = F_0)$. Esto es lo que se llama un **contraste de bondad de ajuste**.

Calculamos el **estadístico de Kolmogorov-Smirnov**: $D_n := \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$. Si H_0 es cierta se cumple que $D_n \rightarrow 0$ c.s. La idea es rechazar H_0 si D_n es *suficientemente grande*.

Parecería complicado determinar cuánto de grande tiene que ser D_n para rechazar puesto que este umbral podría depender de F_0 . Sin embargo, puede probarse que la distribución de D_n bajo H_0 es la misma para cualquier distribución continua F_0 (se dice que D_n es de **distribución libre**). Esta distribución común está tabulada (o puede simularse fácilmente) lo que permite dar una región crítica del nivel de significación deseado.

En R, el comando en el que está implementado este contraste es `ks.test` :

```
# Ejemplo: Kolmogorov-Smirnov
```

```
set.seed(100)
```

```
# H0 verdadera
```

```
n <- 100
```

```
x <- rnorm(n)
```

```
ks.test(x, "pnorm")
```

```
##
```

```
## One-sample Kolmogorov-Smirnov test
```

```
##
```

```
## data: x
```

```
## D = 0.073373, p-value = 0.6546
```

```
## alternative hypothesis: two-sided
```

```
# H0 falsa
```

```
x <- rchisq(n, 2)
```

```
ks.test(x, "pexp")
```

```
##
```

```
## One-sample Kolmogorov-Smirnov test
```

```
##
```

```
## data: x
```

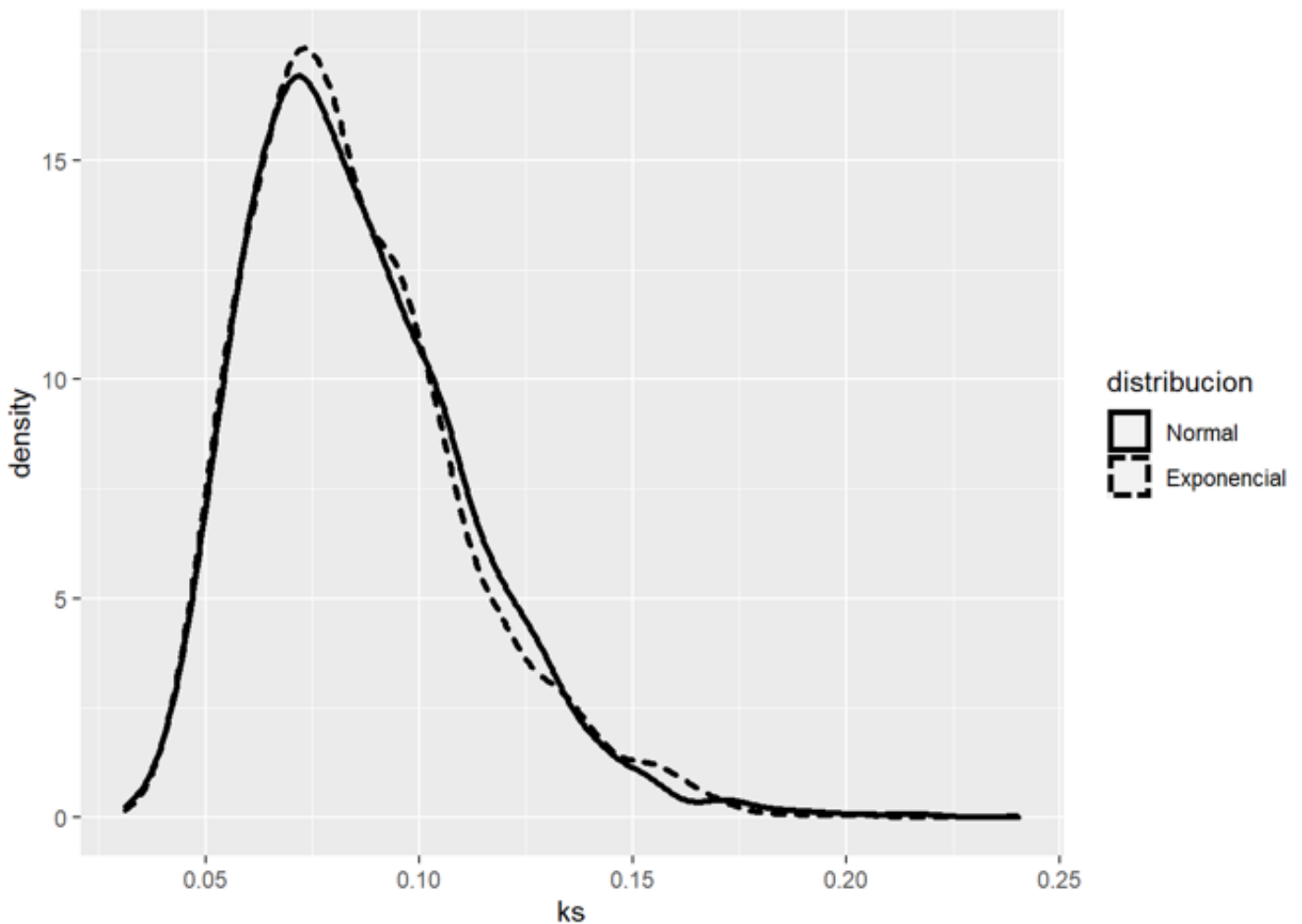
```
## D = 0.25824, p-value = 3.226e-06
```

```
## alternative hypothesis: two-sided
```

El siguiente código simula y representa gráficamente la distribución del estadístico de Kolmogorov-Smirnov. Se hace tanto con datos normales como exponenciales para comprobar que la distribución es la misma en ambos casos.

```
R <- 2000
n <- 100
ks_norm <- replicate(R, ks.test(rnorm(n), "pnorm")$statistic)
ks_exp <- replicate(R, ks.test(rexp(n), "pexp")$statistic)
distribucion <- gl(2, R, labels = c("Normal", "Exponencial"))

df <- data.frame(ks = c(ks_norm, ks_exp), distribucion = distribucion)
ggplot(df) +
  geom_density(aes(x = ks, y = ..density.., linetype = distribucion), size = 1.1)
```

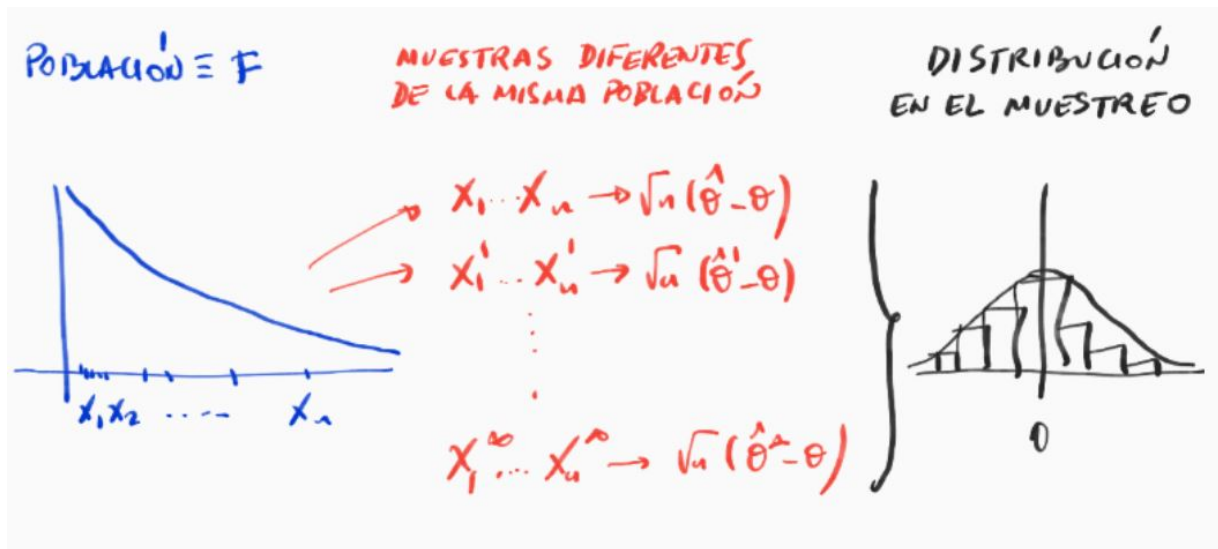


Bootstrap: idea básica

Los procedimientos de remuestreo bootstrap constituyen una metodología general que permite estimar la distribución en el muestreo de una cantidad $T(X_1, \dots, X_n; F)$. Se basan en la combinación de dos ideas:

- El principio de estimación *plug-in* o de sustitución que consiste en sustituir la función de distribución F de la que proceden los datos por la función de distribución empírica F_n en todas aquellas expresiones que dependen de F .
- Un esquema de simulación que permite aproximar numéricamente los estimadores obtenidos según el principio de sustitución. Como veremos, la aplicación de este esquema implica la reutilización de los datos muestrales, lo que justifica el término remuestreo.

Recordamos brevemente el concepto de distribución en el muestreo. Supongamos que $\hat{\theta}$ es un estimador de un parámetro $\theta = \theta(F)$, y que nos interesa la distribución de $T(X_1, \dots, X_n; F) = \sqrt{n}(\hat{\theta} - \theta)$. Esto significa que queremos conocer las probabilidades que toma $\sqrt{n}(\hat{\theta} - \theta)$ para las diferentes muestras posibles. La siguiente figura representa la situación:



Lo que queremos es estimar la distribución de $T(X_1, \dots, X_n; F) = \sqrt{n}(\hat{\theta} - \theta)$ a la derecha de la figura anterior. Matemáticamente, podemos denotar H_n a la función de distribución correspondiente:

$$H_n(x) = P_F(T(X_1, \dots, X_n; F) \leq x),$$

donde la notación P_F indica que la probabilidad se calcula suponiendo que las v.a. X_i tienen distribución F . Esta distribución subyace a la mayoría de técnicas de inferencia (intervalo de confianza, contrastes, etc.) El problema es ser capaces de aproximarla cuando **solo tenemos una** de las infinitas muestras posibles.

Como hemos dicho, el bootstrap esencialmente consiste en sustituir F por F_n , con lo que resulta el estimador bootstrap *ideal*. La aplicación de esta sustitución lleva al estimador

$$\hat{H}_n(x) = P_{F_n}(T(X_1^*, \dots, X_n^*; F_n) \leq x).$$

En la práctica, es necesario poder calcular $\hat{H}_n(x)$ de forma efectiva. Obtener una expresión cerrada suele ser imposible en general, pero al sustituir F por F_n pasamos del duro mundo real, en el que disponemos de una única muestra de tamaño n , al mundo bootstrap en el que la distribución de la que proceden los datos es totalmente conocida, por lo que disponemos de todas las réplicas de $T(X_1^*, \dots, X_n^*; F_n)$ que nuestra capacidad de cálculo permita. En la figura anterior, una vez cambiamos la distribución azul de la izquierda por la empírica, podemos reproducir todo el proceso que aparece en la figura con el ordenador sin necesidad de cálculos matemáticos adicionales.

En la práctica, se simulan muestras bootstrap $X_1^{*b}, \dots, X_n^{*b}$ de F_n , donde $b = 1, \dots, B$ y B es un número muy grande, pero factible. Para cada una de estas muestras artificiales podemos calcular $T^{*(b)} = T(X_1^{*b}, \dots, X_n^{*b}; F_n)$. El valor de $\hat{H}_n(x)$ se puede entonces aproximar de la siguiente forma:

$$\hat{H}_n(x) \approx \frac{1}{B} \sum_{b=1}^B \mathbb{I}_{\{T^{*(b)} \leq x\}}.$$

En resumen, el procedimiento a seguir es:

- Se estima F mediante F_n . (Principio de sustitución o *plug-in*.)
- Se obtienen B muestras bootstrap $X_1^{*b}, \dots, X_n^{*b}$ ($b = 1, \dots, B$) procedentes de la distribución F_n , sorteando con reemplazamiento entre los datos originales X_1, \dots, X_n .
- Se calcula $T^{*(b)} = T(X_1^{*b}, \dots, X_n^{*b}; F_n)$ ($b = 1, \dots, B$) para cada una de las muestras bootstrap.
- Se calcula la proporción $\tilde{H}_B(x) = B^{-1} \sum_{b=1}^B \mathbb{I}_{\{T^{*(b)} \leq x\}}$.

El proceso anterior implica dos aproximaciones diferentes:

$$H_n(x) \approx \hat{H}_n(x) \approx \tilde{H}_B(x).$$

La ley fuerte de los grandes números (aplicada a observaciones de la distribución F_n) garantiza que si $B \rightarrow \infty$ el último término converge al segundo. En esta aproximación todo ocurre dentro del mundo bootstrap. Sin embargo, la aproximación entre los dos primeros términos cuando $n \rightarrow \infty$ requiere trabajo teórico adicional. Es lo que se llama establecer la consistencia del bootstrap. En términos más intuitivos, hay que demostrar que al aumentar el tamaño muestral el mundo bootstrap se parece lo suficiente al mundo real.

El mismo algoritmo sirve esencialmente para estimar cualquier aspecto de la distribución, en lugar de la función de distribución completa. Por ejemplo, muchas veces lo que interesa es estimar la varianza o la desviación típica de un estimador $\hat{\theta}$, es decir, determinar su error típico. En este caso el estimador bootstrap *ideal* de la varianza de $\hat{\theta}$ es $\text{Var}_{F_n}(\hat{\theta}^*)$. En la práctica usaremos la correspondiente aproximación basada en B remuestras:

$$\text{Var}_{F_n}(\hat{\theta}^*) \approx \frac{1}{B-1} \sum_{j=1}^B (\hat{\theta}_j^* - \bar{\theta}^*)^2,$$

donde $\hat{\theta}_j^*$ es el valor del estimador para la remuestra j , y $\bar{\theta}^* = B^{-1} \sum_{j=1}^B \hat{\theta}_j^*$ es el promedio de todas las versiones bootstrap.

Veamos un ejemplo del método que acabamos de describir.

Ejemplo. ¿Cuál es la varianza de la mediana?

Supongamos que X_1, \dots, X_n son v.a.i.i.d. de una distribución de Cauchy (https://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_Cauchy) centrada en θ y con parámetro de escala igual a uno. La función de densidad es:

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad x \in \mathbb{R}.$$

La esperanza de esta distribución no existe, por lo que para estimar θ se usa la mediana. ¿Cuál es la desviación típica de esta mediana? Necesitamos estimar este valor para calcular, por ejemplo, un intervalo de confianza para θ .

En el código siguiente usaremos una muestra original con $n = 30$ datos y $R = 1000$ remuestras.

```
set.seed(100)

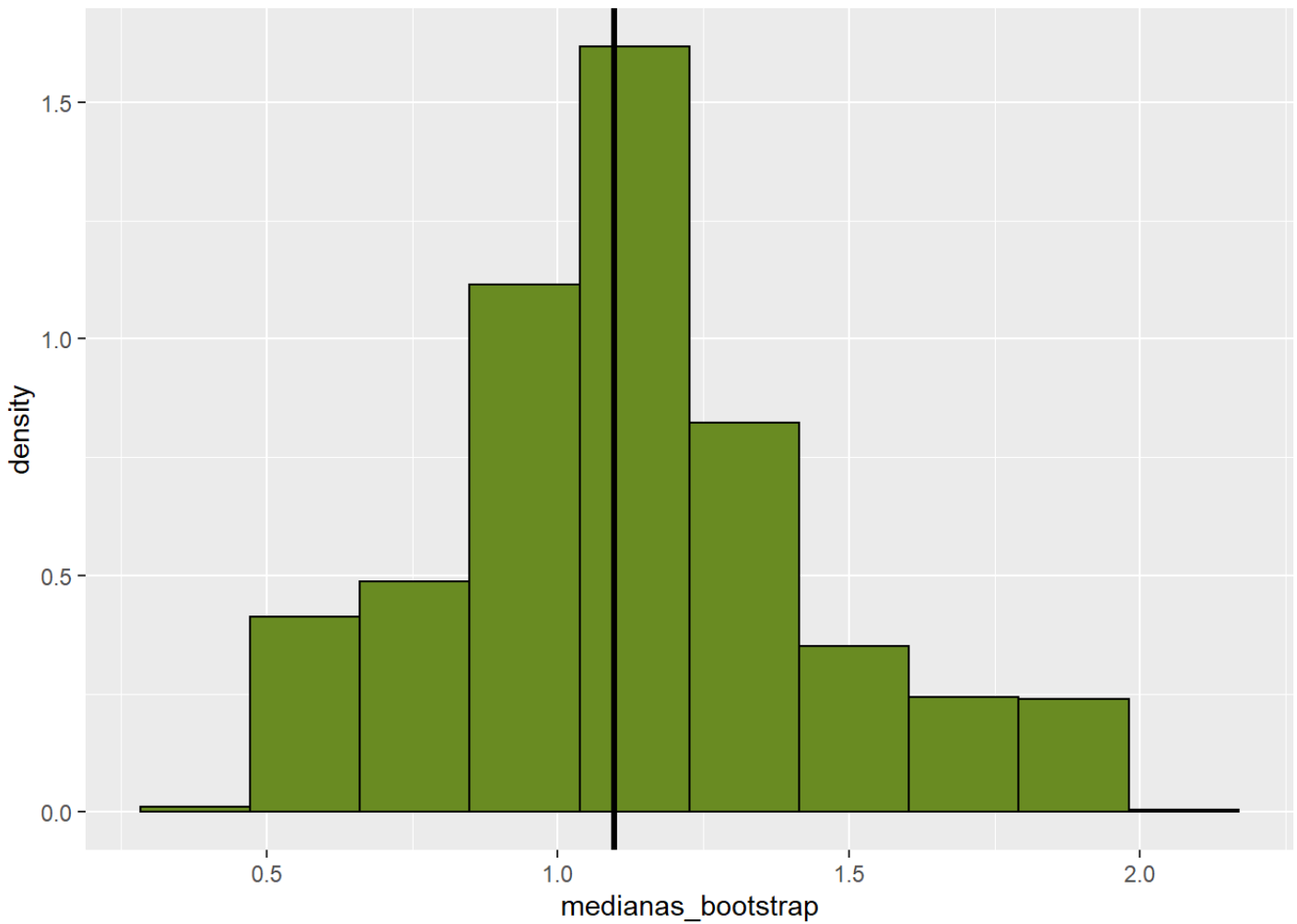
# Parámetros
R <- 1000
n <- 30
theta <- 1

# Generamos los datos
muestra_original <- rt(n, 1) + theta  # Cauchy con theta = 0 coincide con t Student con 1 gl
mediana_original <- median(muestra_original)

# Generamos las remuestras (matriz n x R, cada columna una remuestra)
muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)

# Medianas de las remuestras
medianas_bootstrap <- apply(muestras_bootstrap, 2, median)

# Histograma de las medianas bootstrap
df <- data.frame(medianas_bootstrap = medianas_bootstrap)
ggplot(df) +
  geom_histogram(aes(x = medianas_bootstrap, y = ..density..),
                 bins = 10, fill = 'olivedrab4', col = 'black') +
  geom_vline(xintercept = mediana_original, size = 1.1)
```

```
# Estimador bootstrap de la desviación típica de la mediana
sd_mediana <- sd(medianas_bootstrap)
sd_mediana
```

```
## [1] 0.3135752
```

Los resultados anteriores parecen sugerir que la distribución de la mediana no está lejos de ser normal. Además, el error típico de la mediana obtenido mediante bootstrap ha sido 0.314.

Podemos comparar estos resultados con los que se obtendrían usando la teoría asintótica clásica para la mediana. Si M_n es la mediana muestral, $m = F^{-1}(1/2)$ es la única mediana poblacional, y la distribución de la que proceden las observaciones tiene una función de densidad f , continua en un entorno de m tal que $f(m) > 0$, se verifica

$$\sqrt{n}(M_n - m) \rightarrow_d N\left(0, \frac{1}{4f(m)^2}\right).$$

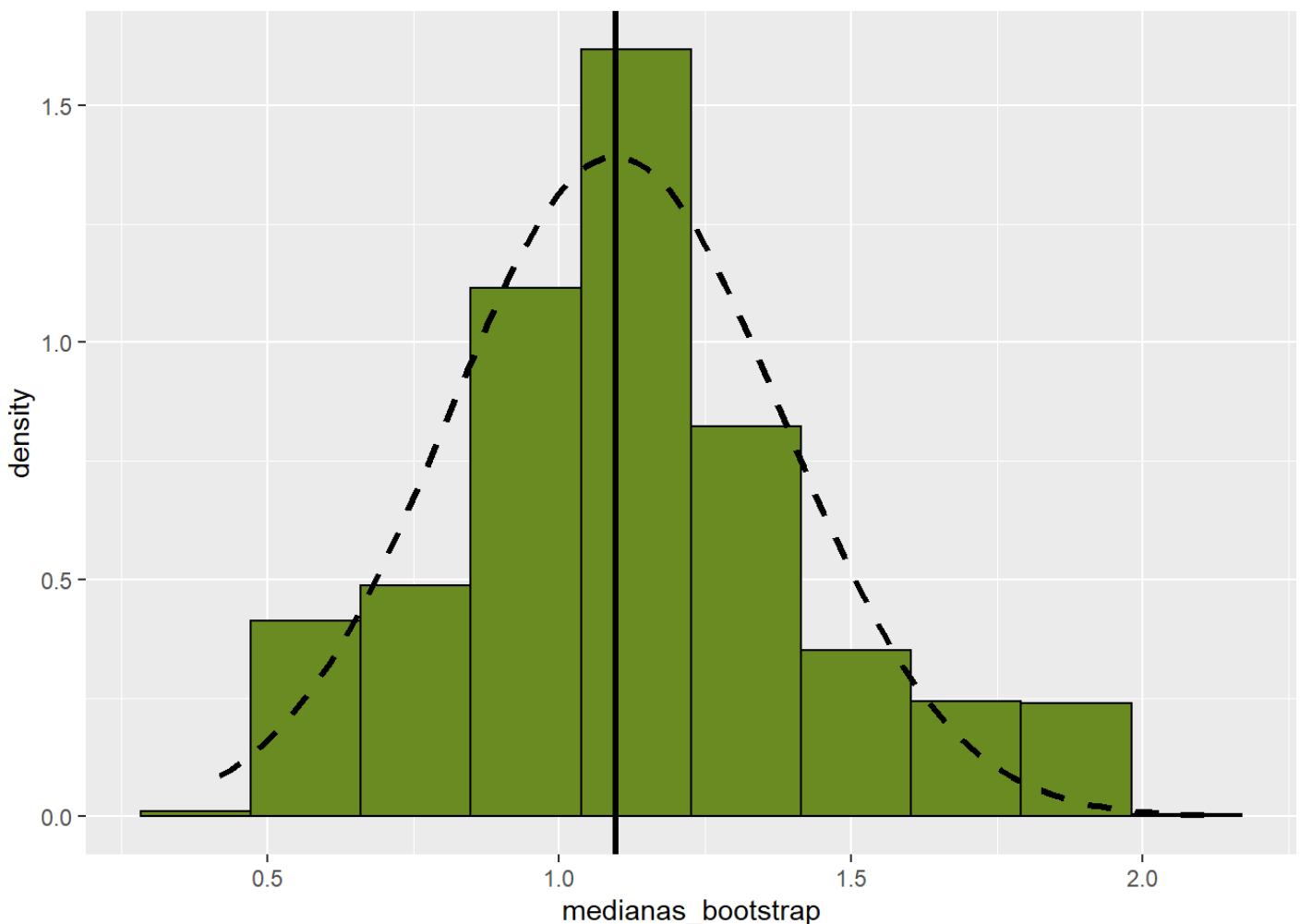
Para n grande podemos esperar

$$M_n \cong N\left(m, \frac{1}{4nf(m)^2}\right)$$

Para la distribución de Cauchy, tenemos $f(m) = 1/\pi$, por lo que su desviación típica si n grande podríamos aproximarla mediante $0.5\pi/\sqrt{n} \approx 0.29$, si $n = 30$. Este valor no está lejos del obtenido mediante bootstrap.

Superponemos a continuación, la distribución normal aproximada al histograma de medianas bootstrap anterior:

```
ggplot(df) +
  geom_histogram(aes(x = medianas_bootstrap, y = ..density..),
                 bins = 10, fill = 'olivedrab4', col = 'black') +
  geom_vline(xintercept = mediana_original, size = 1.1) +
  stat_function(fun = dnorm,
               args = list(mean = mediana_original, sd = .5*pi/sqrt(n)),
               linetype = 2, size = 1.1)
```



Consistencia

Como muestra el ejemplo anterior, el bootstrap nos ha permitido reemplazar un TCL para la mediana, que no es trivial de obtener, por cálculos llevados a cabo con el ordenador. Sin embargo, garantizar la validez asintótica del procedimiento no es tampoco fácil.

Dada una distancia entre distribuciones de probabilidad ρ , se dice que el bootstrap es fuertemente consistente si $\rho(H_n, \hat{H}_n) \rightarrow 0$ con probabilidad 1, si $n \rightarrow \infty$. Es débilmente consistente si $\rho(H_n, \hat{H}_n) \rightarrow_p 0$. En los primeros artículos que incluyeron este tipo de resultados se consideraron la distancia de Kolmogorov

$\rho(F, G) = \|F - G\|_\infty = \sup_x |F(x) - G(x)|$ y la llamada distancia de Mallows (https://en.wikipedia.org/wiki/Wasserstein_metric).

Uno de los primeros resultados de validez asintótica del bootstrap para estimar la distribución de la media muestral es el siguiente:

Teorema (Singh, 1981). (<https://projecteuclid.org/euclid.aos/1176345636>) Supongamos $E_F(X^2) < \infty$ y denotemos $\mu = E_F(X)$, $H_n(x) = P_F(\sqrt{n}(\bar{X} - \mu) \leq x)$ y $\hat{H}_n(x) = P_{F_n}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x)$. Entonces $\|H_n - \hat{H}_n\|_\infty \rightarrow 0$, con probabilidad 1.

Resultados muy generales de validez del bootstrap para la mediana se pueden encontrar en Ghosh et al (1984) (<https://projecteuclid.org/euclid.aos/1176346731>).

Teorema (Ghosh et al., 1984). (<https://projecteuclid.org/euclid.aos/1176346731>) Sea X_1, \dots, X_n una muestra de v.a.i.i.d. de una distribución F tal que $E_F|X|^\alpha < \infty$, para algún $\alpha > 0$. Supongamos que F tiene una única mediana poblacional m y una función de densidad f continua en un entorno de m tal que $f(m) > 0$. Entonces

$$\text{Var}_{F_n}(\sqrt{n}(M_n^* - M_n)) \rightarrow_{c.s.} \frac{1}{4f(m)^2}.$$

Según el resultado anterior, el estimador bootstrap de la varianza converge con probabilidad 1 al valor que obtendríamos usando la distribución asintótica, **pero el bootstrap no requiere conocer la función de densidad f .**

¿Cuándo falla el bootstrap?

El método bootstrap no siempre es consistente. La situación típica en la que falla es aquella en la que el estadístico $T(X_1, \dots, X_n; F)$ no admite un teorema central del límite. Por ejemplo,

- $T(X_1, \dots, X_n; F) = \sqrt{n}(\bar{X} - \mu)$, pero $\text{Var}(X) = \infty$.
- $T(X_1, \dots, X_n; F) = \sqrt{n}(g(\bar{X}) - g(\mu))$, pero g no es derivable en μ .
- $T(X_1, \dots, X_n; F) = \sqrt{n}(F_n^{-1}(p) - F^{-1}(p))$, pero $f(F^{-1}(p)) = 0$.
- La distribución de los datos es F_θ y el soporte de F_θ depende del parámetro.

Vamos a ver un ejemplo de esta última situación. Sea X_1, \dots, X_n va iid de una distribución uniforme en el intervalo $(0, \theta)$. El estimador de máxima verosimilitud de θ es $\hat{\theta} = X_{(n)}$, el máximo de las observaciones X_1, \dots, X_n . Consideramos la distribución asintótica de $T_n = n(\theta - X_{(n)})$. Si $x \geq 0$ y n suficientemente grande de manera que

$$x/n < \theta,$$

$$P_F(T_n \leq x) = 1 - P_F(X_{(n)} \leq \theta - x/n) = 1 - \left(\frac{\theta - x/n}{\theta} \right)^n \rightarrow 1 - e^{-x/\theta},$$

si $n \rightarrow \infty$. Supongamos que $\theta = 1$ y, por lo tanto, $T_n \rightarrow_d \exp(1)$. Consideremos ahora la versión bootstrap $T_n^* = n(X_{(n)} - X_{(n)}^*)$. Para $x \geq 0$ se verifica

$$P_{F_n}(T_n^* \leq x) \geq P_{F_n}(T_n^* = 0) = P_{F_n}(X_{(n)}^* = X_{(n)}) = 1 - \left(\frac{n-1}{n} \right)^n \rightarrow 1 - e^{-1}.$$

Podemos tomar, por ejemplo, $x = 0.001$ para comprobar que $P_F(T_n \leq x)$ y $P_{F_n}(T_n^* \leq x)$ no pueden tener el mismo límite.

Intervalos de confianza

Una vez que hemos estimado la distribución en el muestreo de $\sqrt{n}(\hat{\theta} - \theta)$, podemos usar la estimación para deducir intervalos de confianza para θ . Existe una literatura muy amplia sobre el cálculo de intervalos de confianza mediante bootstrap. Aquí vamos a revisar brevemente uno de los métodos más directos, conocido en la literatura como *método híbrido*.

Si la distribución $H_n(x)$ fuese conocida, entonces se podría obtener un intervalo de confianza para θ de nivel exacto $1 - \alpha$ despejando θ en la ecuación siguiente:

$$1 - \alpha = P_F\{H_n^{-1}(\alpha/2) \leq \sqrt{n}(\hat{\theta} - \theta) \leq H_n^{-1}(1 - \alpha/2)\}.$$

El intervalo de confianza correspondiente sería

$$(\hat{\theta} - n^{-1/2} H_n^{-1}(1 - \alpha/2), \hat{\theta} - n^{-1/2} H_n^{-1}(\alpha/2)).$$

Dado que H_n no es conocida, resulta natural reemplazarla por el estimador bootstrap \hat{H}_n . En la práctica, esto requiere ordenar todos los valores simulados $\sqrt{n}(\hat{\theta}^{*b} - \hat{\theta})$, seleccionar los percentiles que dejan una proporción de valores $\alpha/2$ a su izquierda y a su derecha, y utilizar estos valores en la fórmula anterior en el lugar de $H_n^{-1}(\alpha/2)$ y $H_n^{-1}(1 - \alpha/2)$, respectivamente.

En el siguiente ejemplo se aplica el método bootstrap híbrido para calcular un intervalo de confianza (nivel de confianza nominal $1 - \alpha = 0.95$) para la mediana de una distribución de Cauchy. Se replica el procedimiento $m = 100$ muestras de tamaño $n = 30$ y se determina el número de ellas en las que el intervalo contiene al verdadero valor del parámetro:

```
set.seed(100)

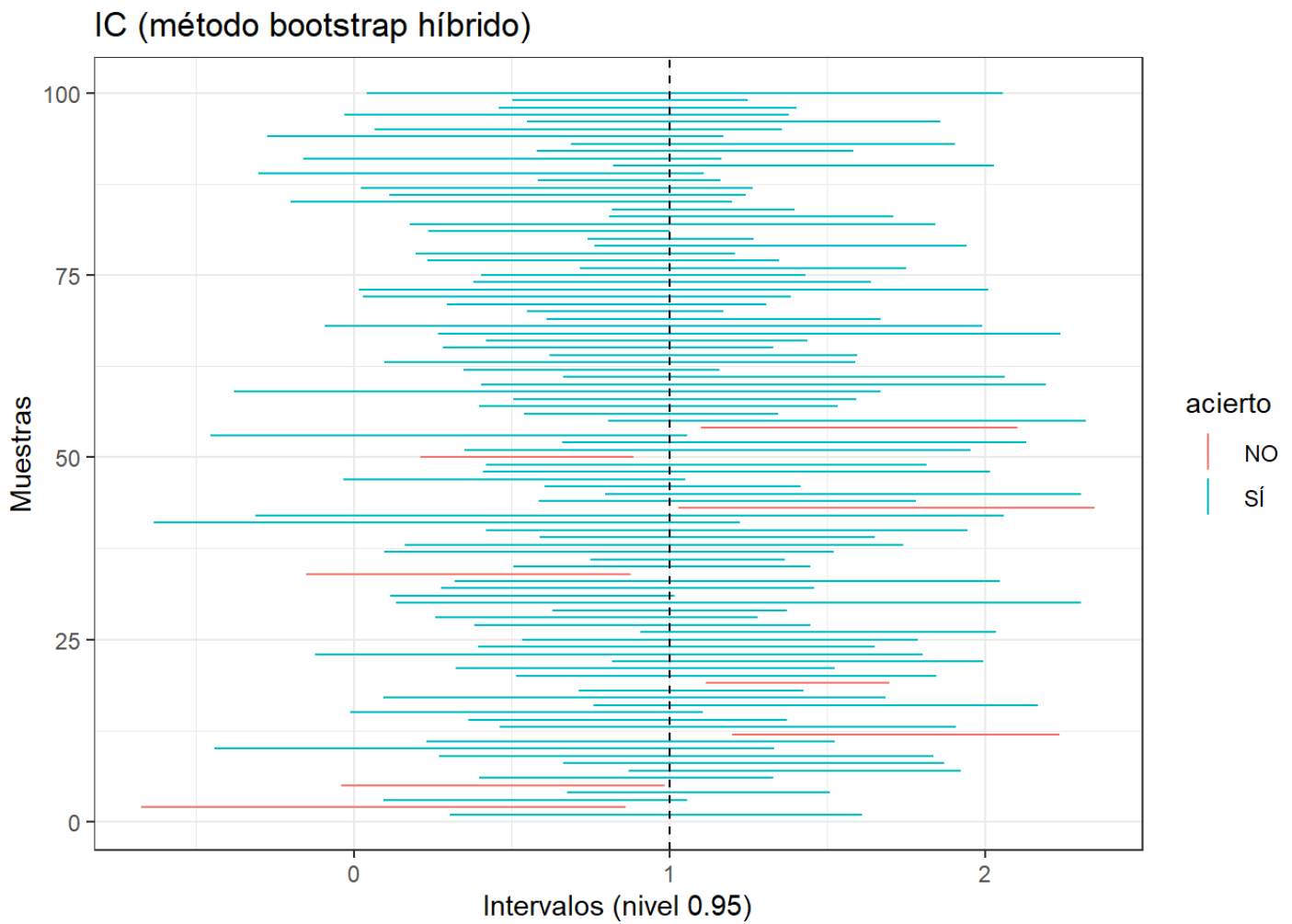
# Parámetros
R <- 1000
n <- 30
theta <- 1
m <- 100
alfa <- 0.05

# Cálculo de los intervalos
acierto <- NULL
intervalo <- NULL
for (i in 1:m){
  muestra_original <- rt(n, 1) + theta
  mediana_original <- median(muestra_original)

  muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)
  medianas_bootstrap <- apply(muestras_bootstrap, 2, median)
  T_bootstrap <- sqrt(n) * (medianas_bootstrap - mediana_original)
  ic_min <- mediana_original - quantile(T_bootstrap, 1-alfa/2)/sqrt(n)
  ic_max <- mediana_original + quantile(T_bootstrap, alfa/2)/sqrt(n)
  intervalo <- rbind(intervalo, c(ic_min, ic_max))
  acierto <- c(acierto, ic_min < theta & ic_max > theta)
}

# Gráfico
df <- data.frame(ic_min <- intervalo[,1],
                 ic_max <- intervalo[, 2],
                 ind = 1:m,
                 acierto = acierto)

ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = acierto)) +
  scale_color_hue(labels = c("NO", "Sí")) +
  geom_vline(aes(xintercept = theta), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'Intervalos (nivel 0.95)',
       title = 'IC (método bootstrap híbrido)')
```



Existen otros métodos más refinados para obtener intervalos de confianza mediante bootstrap. Por ejemplo, otra posibilidad es basar el intervalo bootstrap en una versión estudentizada $(\hat{\theta} - \theta)/\hat{\sigma}_n$, donde $\hat{\sigma}_n$ es un estimador de la desviación típica de $\hat{\theta}$.

Referencias

James et al. (2013) presenta una perspectiva aplicada del bootstrap. Para profundizar en las relaciones del bootstrap con otros métodos se puede consultar Efron y Hastie (2016). DasGupta (2008) contiene un buen resumen de las propiedades de consistencia.

- DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer.
- Efron, B. y Hastie, T. (2016). (<http://web.stanford.edu/~hastie/CASI/>) *Computer Age Statistical Inference*. Cambridge University Press.
- James, G., Witten, D., Hastie, T. y Tibshirani, R. (2013). (<http://faculty.marshall.usc.edu/gareth-james/ISL/>) *An introduction to statistical learning*. Springer.